Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Imaging-based Computational Biomedicine (Yoshinobu Sato (Professor))		
Student ID	1911317	Submission date	2021 / 1 / 25
Name	LIANG TIANHENG		
Thesis title	Unsupervised Classification of Dilated/Hypertrophic Cardiomyopathy in Pathological Images: Quantitative Analysis 拡張型/ 肥大型心筋症病理画像の教師なし分類と評価		

Abstract

The diagnosis of dilated/hypertrophic cardiomyopathy (DCM/HCM) using pathological images requires expertise in the field but the number of expert doctors is limited. Also it is a time-consuming task, highly depends on pathologists' experience and knowledge. Many image-based supervised learning approaches have been proposed to classify pathological images. However, given the patch diversity with respect to cellular contents, the assignment of a ground-truth label for each patch (i.e., patch-wise label), rather than a global (i.e. image-wise) label, is essential as it affects the training reliability. This project aims at leveraging unsupervised learning approaches to quantitatively assess the patch diversity with respect to disease classes, followed by a pathologists' annotation analysis to discuss the reliability of image-wise labeling. The proposed approach uses t-distributed stochastic neighbor embedding (t-SNE) and k-means clustering to assemble the patches into distinct clusters, InceptionV3 ¥cite{szegedy2016rethinking} as supervised learning model to evaluated the results. In this study, 276 images from 96 cardiomyopathy patients were used. The images were split into patches with a matrix size of 256×256. The patches including the background and boundaries were excluded, thus 7423 patches were used in the analysis. The patches were grouped into 300 clusters using t-SNE with a perplexity of 50. Four expert medical doctors were asked to classify the clusters into DCM, HCM or normal classes. The patch diversity was assessed by evaluating the agreement among the experts on the classification of each cluster. The impact of he patch diversity was assessed by using a supervised classification model (InceptionV3). The results have shown that a higher accuracy can be obtained using patch-wise rather than image-wise labels. These results indicate the intrinsic diversity in the patch classifications and the possibility to obtain a more accurate accurate method in building supervised learning data set for pathological images.